

Use of DG008, DG065, DG210 or DG239 secreted protein products for preventing and treating pancreatic diseases and/or obesity and/or metabolic syndrome

Description

This invention relates to the use of secreted DG008, DG065, DG210, or DG239 proteins, to the use of polynucleotides encoding these, and to the use of effectors/modulators thereof in the diagnosis, study, prevention, and treatment of pancreatic diseases (e.g. diabetes mellitus), obesity and/or metabolic syndrome and to the use in regeneration of tissues such as pancreatic tissues and others.

Many human proteins serve as pharmaceutically active compounds. Several classes of human proteins that serve as such active compounds include hormones, cytokines, cell growth factors, and cell differentiation factors. Most proteins that can be used as a pharmaceutically active compound fall within the family of secreted proteins. Secreted proteins are generally produced within cells at rough endoplasmic reticulum, are then exported to the golgi complex, and then move to secretory vesicles or granules, where they are secreted to the exterior of the cell via exocytosis. Examples for commercially used secreted proteins are human insulin, thrombolytic agents, interferons, interleukins, colony stimulating factors, human growth hormone, transforming growth factor beta, tissue plasminogen activator, erythropoietin, and various other proteins. Receptors of secreted proteins, which are membrane-bound proteins, also have potential as therapeutic or diagnostic agents. It is, therefore, important for developing new pharmaceutical compounds to identify secreted proteins that can be tested for activity in a variety of animal models. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel functions for human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical diseases, disorders, and/or conditions by using secreted proteins or the genes that encode them.

The pancreas is an essential organ possessing both an exocrine function involved in the delivery of enzymes into the digestive tract and an endocrine function by which various hormones are secreted into the blood stream. The

exocrine function is assured by acinar and centroacinar cells that produce various digestive enzymes and intercalated ducts that transport these enzymes in alkaline solution to the duodenum. The functional unit of the endocrine pancreas is the islet of Langerhans. Islets are scattered throughout the exocrine portion of the pancreas and are composed of four cell types: alpha-, beta-, delta- and PP-cells, reviewed for example in Kim S.K. and Hebrok M., (2001) *Genes Dev.* 15: 111-127. Beta-cells produce insulin, represent the majority of the endocrine cells and form the core of the islets, while alpha-cells secrete glucagon and are located in the periphery. Delta-cells and PP-cells are less numerous and secrete somatostatin and pancreatic polypeptide, respectively.

Early pancreatic development has been well studied in different species, including chicken, zebrafish, and mice (for a detailed review, see Kim & Hebrok, 2001, *supra*). The pancreas develops from distinct dorsal and ventral structures. Pancreas development requires specification of the pancreas structure along both anterior-posterior and dorsal-ventral axes. A number of transcription factors, which are critical for proper pancreatic development have been identified (see Kim & Hebrok, 2001, *supra*; Wilson M.E. et al., (2003) *Mech Dev.* 120: 65-80).

In postnatal/adult humans, the acinar and ductal cells retain a significant proliferative capacity that can ensure cell renewal and growth, whereas the islet cells become mostly mitotically inactive. This is in contrast to rodents where β -cell replication is an important mechanism in the generation of new beta cells. It has been suggested, that during embryonic development, pancreatic islets of Langerhans originate from differentiating duct cells or other cells with epithelial morphology (Bonner-Weir S. and Sharma A., (2002) *J Pathol.* 197: 519-526; Gu G. et al., (2003) *Mech Dev.* 120: 35-43). In adult humans, new beta cells arise in the vicinity of ducts (Butler A.E. et al., (2003) *Diabetes* 52: 102-110; Bouwens L. and Pipeleers D.G., (1998) *Diabetologia* 41: 629-633). However, also an intra-islet location or an origin in the bone marrow has been suggested for precursor cells of adult beta cells (Zulewski H. et al., (2001) *Diabetes* 50: 521-533; Ianus A. et al., (2003) *J Clin Invest.* 111: 843-850). Pancreatic islet growth is dynamic and responds to changes in insulin demand, such as during pregnancy or during the increase in body mass occurring during childhood. In adults, there is a good correlation between body

mass and islet mass (Yoon K.H. et al., (2003) J Clin Endocrinol Metab. 88: 2300-2308).

- Pancreatic beta-cells secrete insulin, which is stimulated by high blood glucose levels. Insulin amongst other hormones plays a key role in the regulation of the fuel metabolism. Insulin leads to the storage of glycogen and triglycerides and to the synthesis of proteins. The entry of glucose into muscles and adipose cells is stimulated by insulin. In patients who suffer from diabetes mellitus the amount of insulin produced by the pancreatic islet cells is too low, resulting in elevated blood glucose levels (hyperglycemia). In diabetes type 1, beta-cells are lost due to autoimmune destruction. In type 2 diabetic patients, liver and muscle cells lose their ability to respond to normal blood insulin levels (insulin resistance). High blood glucose levels (and also high blood lipid levels) lead to an impairment of beta-cell function and to an increase in beta-cell apoptosis. It is interesting to note that the rate of beta-cell neogenesis does not appear to change in type 2 diabetics (Butler et al., 2003 supra), thus causing a reduction in total beta-cell mass over time. Eventually the application of exogenous insulin becomes necessary in type 2 diabetics.
- Improving metabolic parameters such as blood sugar and blood lipid levels (e.g. through dietary changes, exercise, medication or combinations thereof) before beta cell mass has fallen below a critical threshold leads to a relatively rapid restoration of beta cell function. However, after such a treatment the pancreatic endocrine function would remain impaired due to the only slightly increased regeneration rate.

In type 1 diabetics, the lifespan of pancreatic islets is dramatically shortened due to autoimmune destruction. Treatments have been devised which modulate the immune system and may be able to stop or strongly reduce islet destruction (Raz I. et al., (2001) Lancet 358: 1749-1753; Chatenoud L. et al., (2003) Nat Rev Immunol. 3: 123-132). However, due to the relatively slow regeneration of human beta cells such treatments could only be fully successful at improving the diabetic condition if they are combined with an agent which can stimulate beta cell regeneration.

Thus, both for type 1 and type 2 diabetes (early and late stages) there is a need to find novel agents which stimulate beta cell regeneration.

4

Diabetes is a very disabling disease, because medications do not control blood sugar levels well enough to prevent swinging between high and low blood sugar levels. Patients with diabetes are at risk for major complications, including diabetic ketoacidosis, end-stage renal disease, diabetic retinopathy and amputation. There are also a host of related conditions, such as metabolic syndrome, obesity, hypertension, heart disease, peripheral vascular disease, and infections, for which persons with diabetes are at substantially increased risk. The treatment of these complications contributes to a considerable degree to the enormous cost which is imposed by diabetes on health care systems world wide.

Obesity is one of the most prevalent metabolic disorders in the world. It is still a poorly understood human disease that becomes as a major health problem more and more relevant for western society. Obesity is defined as a body weight more than 20% in excess of the ideal body weight, frequently resulting in a significant impairment of health. Obesity may be measured by body mass index, an indicator of adiposity or fatness. Further parameters for defining obesity are waist circumferences, skinfold thickness and bioimpedance. It is associated with an increased risk for cardiovascular disease, hypertension, diabetes mellitus type II, hyperlipidaemia and an increased mortality rate. Obesity is influenced by genetic, metabolic, biochemical, psychological, and behavioral factors and can be caused by different reasons such as non-insulin dependent diabetes, increase in triglycerides, increase in carbohydrate bound energy and low energy expenditure (Kopelman P.G., (2000) Nature 404: 25 635-643).

The concept of 'metabolic syndrome' (syndrome x, insulin-resistance syndrome, deadly quartet) was first described 1966 by Camus and reintroduced 1988 by Reaven (Camus J.P., (1966) Rev Rhum Mal Osteoartic 30 33: 10-14; Reaven G.M., (1988), Diabetes 37: 1595-1607). Today, metabolic syndrome is commonly defined as clustering of cardiovascular risk factors like hypertension, abdominal obesity, high blood levels of triglycerides and fasting glucose as well as low blood levels of HDL cholesterol. Insulin resistance greatly increases the risk of developing the metabolic syndrome (Reaven G., (2002) Circulation 106: 286-288). The metabolic syndrome often precedes the development of type II diabetes and cardiovascular disease (Lakka H.M. et al., (2002) JAMA 288: 2709-2716). The control of blood lipid levels and blood

glucose levels is essential for the treatment of the metabolic syndrome (see, for example, Santomauro A.T. et al., (1999) *Diabetes*, 48: 1836-1841).

The molecular factors regulating food intake and body weight balance are incompletely understood. Even if several candidate genes have been described which are supposed to influence the homeostatic system(s) that regulate body mass/weight, like leptin or the peroxisome proliferator-activated receptor-gamma co-activator, the distinct molecular mechanisms and/or molecules influencing obesity or body weight/body mass regulations are not known.

There is a need in the prior art for the identification of candidate genes that are specifically expressed in early development in certain pancreatic tissues. These genes and the thereby encoded proteins can provide tools to the diagnosis and treatment of severe pancreatic disorders and related diseases. Therefore, this invention describes secreted proteins that are specifically expressed in pancreatic tissues early in the development. The invention relates to the use of these genes and proteins in the diagnosis, prevention and/or treatment of pancreatic dysfunctions, such as diabetes, and other related diseases such as obesity and/or metabolic syndrome. These proteins and genes are especially useful in regeneration processes, such as regeneration of the pancreas cells.

In this invention, we disclose secreted factors referred to as DG008, DG065, DG210, and DG239, which are involved in pancreas development, regeneration, and in the regulation of energy homeostasis.

DG008 corresponds to human Sparc-like 1 protein (SPARCL1, hevin). SPARCL1 is a member of the SPARC family of extracellular matrix proteins (Girard J.P. and Springer T.A., (1995) *Immunity* 2: 113-123). SPARCL1 is secreted acidic calcium-binding glycoprotein. The protein was shown to be a negative regulator of cell proliferation. More specifically, it was shown that SPARCL1 can inhibit progression of cells from G1 to S phase or prolong G1 phase (Claeskens A. et al., (2000) *Br J Cancer* 82: 1123-1130). In the extracellular matrix of cultured osteosarcoma cells SPARCL1 was found associated with collagen I-containing fibrils (Hambrock H.O. et al., (2003) *J Biol Chem*. 278: 11351-11358) and it was shown to be involved in metastatic

prostate adenocarcinoma (Nelson P.S., (1998) *Cancer Res.* 58: 232-236). Several uses of SPARCL1 in the diagnosis and treatment of cancers, such as, for example, breast cancer (WO 02/10436), ovarian cancer (WO 02/102235), lung cancer (WO 00/55180, WO 03/029273), metastatic colorectal cancer (WO 02/068677), prostate cancer (WO 03/012067), and leukemia (WO 01/57190) were suggested. In addition, SPARCL1 could be used in diagnosing, treating or preventing diseases associated with matrix remodelling (WO 00/21986).

DG065 corresponds to human SPARC related modular calcium binding 2 (SMOC2). The mRNA encoding this secreted modular protein was described to contain an EF-hand calcium-binding domain, two thyroglobulin-like domains, a follistatin-like domain and a novel domain found only in the homologous SMOC1 (see Vannahme C. et al., (2003) *Biochem J.* 373: 805-814). It was shown by analysis of the recombinantly expressed protein that SMOC2 is a glycoprotein with a calcium-dependent conformation. Further, SMOC2 showed a widespread expression in many tissues. (Vannahme C. et al., *supra*). No further functional data are available in the scientific prior art for this protein.

DG210 corresponds to human secreted frizzled-related protein 1 (SFRP1), a member of the SFRP family that contains a cysteine-rich domain homologous to the putative Wnt-binding site of Frizzled proteins (Rattner A. et al., (1997) *Proc Natl Acad Sci* 94: 2859-2863; Finch P.W. et al., (1997) *Proc. Natl. Acad. Sci.* 94: 6770-6777). SFRPs act as soluble modulators of Wnt signaling. SFRP1 may be involved in determining the polarity of photoreceptor cells in the retina. SFRP1 is expressed in several human tissues, with the highest expression levels in heart.

A function for SFRP1 as tumor suppressor has been postulated (see, for example, WO 98/54325). For example, it is down-regulated and induces apoptosis in normal cervical epithelium and down-regulation of hsFRP contributes to development of cervical cancer (Ko J. et al., (2002) *Exp. Cell Res.* 280: 280-287). A role in the development of other cancers was postulated, for example, breast cancer (WO 02/059377, WO 00/55629), metastatic colorectal cancer (WO 02/068677), and bladder cancer (WO 03/003906). Also, a method for diagnosing and treating glaucomas comprising

analysing and modulating the expression and activity of SFRP1 gene and a gene of the Wnt signal transduction pathway was suggested because an aberrant level of expression of a Wnt pathway component or SFRP1 indicates the presence of or a predisposition to a glaucoma (see, WO 01/64949). In
5 addition, pharmaceutical compositions and methods of use in regulation of mammalian bone forming activities of SFRPs were described in WO 01/19855. SFRP1 was shown to be regulated by osteogenic agents in a differentiation selective manner modulating the life of osteoblasts/preosteocytes.

10 DG239 corresponds to human granulin precursor, a member of the granulin/epithelin growth factor family of proteins. The granulin/epithelin motif defines a family of structurally unique proteins and of great evolutionary antiquity. Granulins are 6-kD peptides with growth modulatory effects on a variety of cells. The granulin/epithelin precursor gene is expressed
15 ubiquitously, but its expression is predominantly in epithelial and haematopoietic cells (for review, see Bateman A. and Bennett H.P., (1998) J Endocrinol. 158: 145-151). There is a great deal of versatility in the means by which cells process and handle the granulin/epithelin precursor. In some instances, the precursor is secreted intact, and in others it is stored in a
20 vesicular organelle, such as the sperm acrosome (Baba T. et al., (1993) Mol Reprod Dev. 34: 233-243). It may be processed into small 6-kDa peptides, which, in the neutrophil, can also be stored in vesicles (Bateman A. et al., (1990) Biochem Biophys Res Commun. 173: 1161-1168, Couto et al., (1992) Infect Immun. 60: 3065-3071). The 6-kDa peptide, the intact precursor, and
25 related proteins regulate the growth of epithelial and mesenchymal cells.

Accordingly, granulin is known as a growth factor involved in tumorigenesis and development. For example, it was described as growth factor for ovarian cancer, gastric cancer, brain tumors, breast cancer, benign prostatic hyperplasia, and others (see, for example, WO 91/15510; WO 00/36107;
30 Jones M.B. et al., (2003) Gynecol Oncol. 88: S136-139; Line A. et al., (2002) Br J Cancer. 86: 1824-1830; Markert J.M. et al., (2001) Physiol Genomics. 5: 21-33; Lu R. and Serrero G., (2001) Proc Natl Acad Sci 98: 142-147; WO 00/55350; WO 02/12440), for diagnosing tumourigenicity in a human
35 (US2002025543-A1), as a probable wound-related growth factor for wound response (WO 91/15510; He Z. et al., (2003) Nat Med. 2003 9: 225-229), having a role in host defense and wound repair (WO 93/15195; Zhu J. et al.,

(2002) Cell 111: 867-878), used for diagnosing osteoporosis/bone disease (WO 02/81745), and used to treat psoriasis (WO 91/15510), and for treating hypoxia-regulated conditions (WO 02/46465).

5 Accordingly, the present invention relates to secreted proteins with novel functions in the human metabolism, regeneration, and pancreatic developmental processes. The present invention discloses specific genes and proteins encoded thereby and effectors/modulators thereof involved in the regulation of pancreatic function and/or metabolism, especially in pancreas
10 diseases such as diabetes mellitus, e.g. insulin dependent diabetes mellitus and/or non insulin dependent diabetes mellitus, and/or metabolic syndrome, obesity, and/or related disorders such as coronary heart disease, eating disorder, cachexia, hypertension, hypercholesterolemia (dyslipidemia), liver fibrosis, and/or gallstones. Further, the present invention discloses specific
15 genes and proteins encoded thereby and effectors/modulators thereof involved in the regeneration of pancreatic cells or tissues, e.g. cells having exocrinous functions such as acinar cells, centroacinar cells and/or ductal cells and/or cells having endocrinous functions, particularly cells in Langerhans islets such as alpha-, beta-, delta- and/or PP-cells, more particularly beta-cells.

20 In this invention, we used a screen for secreted factors expressed in developing mammalian (mouse) pancreas, as described in more detail in the Examples section (see Example 1). This screen identified DG008, DG065, DG210, and DG239 as secreted factors expressed in developing mouse
25 pancreas. The present invention describes mammalian DG008, DG065, DG210, and DG239 proteins and the polynucleotides encoding these, in particular human DG008, DG065, DG210, and DG239, as being involved in the conditions and processes mentioned above.

30 The present invention relates to DG008, DG065, DG210, and DG239 polynucleotides encoding polypeptides with novel functions in the development and regeneration of pancreatic tissues and thus in mammalian pancreatic diseases (e.g. diabetes), and also in body-weight regulation, energy homeostasis, and obesity, fragments of said polynucleotides, polypeptides encoded by said polynucleotides or fragments thereof. The invention also
35 relates to vectors, host cells, and recombinant methods for producing the polypeptides and polynucleotides of the invention. The invention also relates

to effectors/modulators of DG008, DG065, DG210, or DG239 polynucleotides and/or polypeptides, e.g. antibodies, biologically active nucleic acids, such as antisense molecules, RNAi molecules or ribozymes, aptamers, peptides or low-molecular weight organic compounds recognizing said polynucleotides or
5 polypeptides.

DG008, DG065, DG210, or DG239 homologous proteins and nucleic acid molecules coding therefore are obtainable from vertebrate species. Particularly preferred are nucleic acids encoding the human DG008, DG065, DG210, or
10 DG239 protein and variants thereof. The invention particularly relates to a nucleic acid molecule encoding a polypeptide contributing to regulating the energy homeostasis and the mammalian metabolism, wherein said nucleic acid molecule comprises

- (a) the nucleotide sequence of human DG008 (SEQ ID NO: 1), DG065
15 (SEQ ID NO: 3) DG210 (SEQ ID NO: 5), and/or DG239 (SEQ ID NO: 7) and/or a sequence complementary thereto,
- (b) a nucleotide sequence which hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a sequence of (a),
- (c) a sequence corresponding to the sequences of (a) or (b) within the
20 degeneration of the genetic code,
- (d) a sequence which encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to the amino acid sequences of the human DG008 (SEQ ID NO: 2), DG065 (SEQ ID NO: 4), DG210
25 (SEQ ID NO: 6), and/or DG239 (SEQ ID NO: 8),
- (e) a sequence which differs from the nucleic acid molecule of (a) to (d) by mutation and wherein said mutation causes an alteration, deletion, duplication and/or premature stop in the encoded polypeptide or
- (f) a partial sequence of any of the nucleotide sequences of (a) to (e)
30 having a length of 15-25 bases, preferably 25-35 bases, more preferably 35-50 bases and most preferably at least 50 bases.

The function of the mammalian DG008, DG065, DG210, or DG239 in
35 mammalian metabolism was validated by analyzing the expression of the transcripts in different tissues and by analyzing the role in adipocyte differentiation (see Example 3 for more detail).

Expression profiling studies (see Examples for more detail) confirm the particular relevance of DG008, DG065, DG210, and DG239 as regulators of energy metabolism in mammals.

- 5 Expression profile (Taqman) analysis revealed that DG008 is expressed in several mammalian tissues, with highest expression levels in hypothalamus and brain in wild type mice. In addition, DG008 is also highly expressed in metabolic active tissue such as white adipose tissue (WAT) and brown adipose tissue (BAT) as depicted in Fig. 2A.

10

Taqman analysis also revealed that DG065 is expressed in several mammalian tissues, with highest expression levels in muscle and spleen of wild type mice. In addition, DG065 is also highly expressed in heart and metabolic active tissues such as WAT, BAT, and kidney as depicted in Fig. 4A.

15

Taqman analysis revealed that DG210 is expressed in several mammalian tissues, with highest expression levels in kidney and WAT in wild type mice. In addition, DG210 is also expressed in BAT, muscle, brain and hypothalamus as depicted in Fig. 6A.

20

Taqman analysis also revealed that DG239 is expressed in several mammalian tissues, with highest expression levels in WAT of wild type mice. In addition, DG239 is also highly expressed in further metabolic active tissues such as BAT and liver, and at lower but robust levels in muscle as depicted in Fig. 8A.

25

Brown adipose tissue (BAT) is a well-characterized tissue, which is well developed in newborn mammals, including humans. One important task of BAT is to generate heat and maintain body temperature homeostasis in newborn. Thus, an expression of DG008, DG065, DG210, and DG239 proteins in adipose tissues is confirming a role in the regulation of metabolism, particularly energy homeostasis and thermogenesis.

30

In mouse models of insulin resistance and/or diabetes, we found, for example, that the expression of DG008 is up-regulated in metabolic active tissue (BAT) which is supporting an essential role of DG008 in the regulation of the mammalian metabolism, particularly in processes related to, obesity, diabetes,

or metabolic syndrome (Fig. 2B).

DG065 is significantly up-regulated in metabolic active tissues (WAT and BAT) of genetically induced obese mice (ob/ob) compared to wild type mice (see Fig. 4B). These data further support an essential role of DG065 in the regulation of the mammalian metabolism, particularly in processes related to obesity, diabetes, or metabolic syndrome. We also found that the expression of DG065 is significantly down-regulated in BAT of starved wild type mice compared to wild type mice fed a standard diet (see Fig. 4B).

10

In mouse models of insulin resistance and/or diabetes, expression of DG210 is up-regulated in metabolic active tissues (BAT, muscle, liver) which is supporting an essential role of DG210 in the regulation of the mammalian metabolism, particularly in processes related to obesity, diabetes, or metabolic syndrome (Fig 6B). In addition, expression of DG210 mRNA is down-regulated in the hypothalamus and small intestine of starved wild type mice compared to wild type mice fed a standard diet (Fig. 6B).

15

In mouse models of insulin resistance and/or diabetes expression, DG239 is significantly up-regulated in metabolic active tissues (BAT) and muscle which is supporting an essential role of DG239 in the regulation of the mammalian metabolism, particularly in processes related to, obesity, diabetes, or metabolic syndrome (Fig. 8B).

20

Expression of DG008, DG065, and DG239 mRNA was also examined in susceptible wild type mice (for example, C57Bl/6) that show symptoms of diabetes, lipid accumulation, and high plasma lipid levels, if fed a high fat diet. We found that expression of DG008 mRNA is up-regulated in WAT in mice showing symptoms of diabetes, lipid accumulation, and high plasma lipid levels, if fed a high fat diet (Fig. 2C). In those mice, the expression of DG065 is down-regulated in kidney supporting that DG065 is involved in the regulation of mammalian metabolism (see Fig. 4C). In addition, expression of DG239 mRNA is significantly up-regulated in WAT in mice showing symptoms of diabetes, lipid accumulation, and high plasma lipid levels, if fed a high fat diet (Fig. 8C). These results demonstrate that the DG008, DG065 and DG239 proteins are playing a role in the regulation of metabolism, particularly energy homeostasis and thermogenesis.

25

30

35

Further, expression of DG008, DG065, DG210, and DG239 mRNA was examined non-obese-diabetic (NOD) mice.

5 In those mice, expression of DG008 is up-regulated in metabolic active tissues (such as liver) and kidney, as shown in Fig. 2D.

The expression of DG065 is significantly up-regulated in BAT compared to wild type mice, supporting that DG065 is involved in the regulation of mammalian metabolism, particularly in energy homeostasis and thermogenesis (see Fig. 10 4D). In addition, expression of DG210 is significantly up-regulated in BAT, colon, brain and liver and down-regulated in small intestine mice (Fig. 6C). Expression of DG239 is down-regulated in pancreas compared to wild-type mice, as shown in Fig. 8D. These results further demonstrate that the DG008, 15 DG065, DG210, and DG239 proteins are playing a role in the regulation of metabolism, particularly energy homeostasis and thermogenesis.

Furthermore, it is shown (see Examples and Fig. 2E) that the DG008 protein is expressed and regulated during the differentiation of the preadipocytes into mature adipocytes. We also show (see Examples and Fig. 4E) that the 20 DG065 mRNA is regulated during differentiation of preadipocytes into mature adipocytes.

Further, we show (see Examples and Figure 8E) that the DG239 mRNA is increased during differentiation of preadipocytes into mature adipocytes.

25 The results are demonstrating a role of DG008, DG065, and DG239 as modulators of adipogenesis.

Thus, DG008, DG065, DG210 and DG239 are strong candidates for the 30 manufacture of pharmaceutical compositions and medicaments for the treatment of conditions related to human metabolism, such as diabetes, obesity, and/or metabolic syndrome.

Microarrays are analytical tools routinely used in bioanalysis. A microarray has 35 molecules distributed over, and stably associated with, the surface of a solid support. The term "microarray" refers to an arrangement of a plurality of

polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as monitoring gene expression, drug discovery, gene sequencing, gene mapping, 5 bacterial identification, and combinatorial chemistry. One area in particular in which microarrays find use is in gene expression analysis (see Example 4). Array technology can be used to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays 10 are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, 15 condition, disease, or disorder.

Microarrays may be prepared, used, and analyzed using methods known in the art (see for example, Brennan T.M., (1995) U.S. Patent No. 5,474,796; Schena M. et al., (1996) Proc. Natl. Acad. Sci. 93: 10614-10619; Baldeschwieler J.D. 20 et al., (1995) PCT application WO 95/251116; Shalon T.D. and Brown P.O., (1995) PCT application WO 95/35505; Heller R.A. et al., (1997) Proc. Natl. Acad. Sci. 94: 2150-2155; Heller, M.J. and Tu E., (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena M., ed. (1999; DNA Microarrays: A Practical Approach, 25 Oxford University Press, London).

Oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques, which monitor the relative 30 expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene 35 expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate

and effective treatment regimen for that patient. For example, therapeutic agents, which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

5 As determined by microarray analysis, DG065 shows differential expression in mouse 3T3-L1 cells and human primary adipocytes. A strong up-regulation is observed concerning the expression of DG065 during adipocyte differentiation (see Fig. 4F and 4G). The DG065 protein in preadipocytes has the potential to enhance adipocyte differentiation. These results demonstrate that the
10 GD065 protein plays an essential role in the regulation of human metabolism, for example, as effector/modulator (for example, enhancer) of adipogenesis. Thus, DG065 is a strong candidate for the manufacture of a pharmaceutical composition and a medicament for the treatment of conditions related to human metabolism, such as diabetes, obesity, and/or metabolic syndrome.

15 The invention also encompasses novel use of polynucleotides that encode the proteins of the invention and homologous proteins. Accordingly, any nucleic acid sequence, which encodes the amino acid sequences of the proteins of the invention and homologous proteins, can be used to generate recombinant molecules that express the proteins of the invention and homologous proteins.
20 In a particular embodiment, the invention encompasses a nucleic acid encoding DG008, DG065, DG210, or DG239. It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding the proteins, some bearing
25 minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. The invention contemplates each and every possible variation of nucleotide sequence that can be made by selecting combinations based on possible codon choices.

30 Also encompassed by the invention is the use of polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those of the polynucleotides encoding the proteins of the invention, under various conditions of stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe, as
35 described in Wahl G.M. et al., (1987; Methods Enzymol. 152: 399-407) and Kimmel A.R. (1987; Methods Enzymol. 152: 507-511), and may be used at a defined stringency. Preferably, hybridization under stringent conditions means

that after washing for 1 h with 1 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 65°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 65°C, a positive hybridization signal is observed.

5 Altered nucleic acid sequences encoding the proteins which are encompassed by the invention include deletions, insertions or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent protein.

10 The encoded proteins may also contain deletions, insertions or substitutions of amino acid residues, which produce a silent change and result in functionally equivalent proteins. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity 15 of the protein is retained. Furthermore, the invention relates to peptide fragments of the proteins or derivatives thereof such as cyclic peptides, retro-inverso peptides or peptide mimetics having a length of at least 4, preferably at least 6 and up to 50 amino acids.

20 Also included within the scope of the present invention are alleles of the genes encoding the proteins of the invention and homologous proteins. As used herein, an 'allele' or 'allelic sequence' is an alternative form of the gene, which may result from at least one mutation in the nucleic acid sequence. Alleles may 25 result in altered mRNAs or polypeptides whose structures or function may or may not be altered. Any given gene may have none, one or many allelic forms. Common mutational changes, which give rise to alleles, are generally ascribed to natural deletions, additions or substitutions of nucleotides. Each of these types of changes may occur alone or in combination with the others, one or more times in a given sequence.

30 The nucleic acid sequences encoding DG008, DG065, DG210, or DG239 and homologous proteins may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements.

35 In order to express a biologically active protein, the nucleotide sequences encoding the proteins or functional equivalents, may be inserted into

appropriate expression vectors, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding the proteins and the appropriate transcriptional and translational control elements. Regulatory elements include for example a promoter, an initiation codon, a stop codon, a mRNA stability regulatory element, and a polyadenylation signal. Expression of a polynucleotide can be assured by (i) constitutive promoters such as the Cytomegalovirus (CMV) promoter/enhancer region, (ii) tissue specific promoters such as the insulin promoter (see, Soria B. et al., (2000), Diabetes 49: 157-162), SOX2 gene promoter (see Li M. et al., (1998) Curr. Biol. 8: 971-974), Msi-1 promoter (see Sakakibara S. and Okano H., (1997) J. Neuroscience 17: 8300-8312), alpha-cardia myosin heavy chain promoter or human atrial natriuretic factor promoter (Klug M.G. et al., (1996) J. Clin. Invest 98: 216-224; Wu J. et al., (1989) J. Biol. Chem. 264: 6472-6479) or (iii) inducible promoters such as the tetracycline inducible system. Expression vectors can also contain a selection agent or marker gene that confers antibiotic resistance such as the neomycin, hygromycin or puromycin resistance genes. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook J. et al., (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y. and Ausubel F.M. et al., (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

In a further embodiment of the invention, natural, modified or recombinant nucleic acid sequences encoding the proteins of the invention and homologous proteins may be ligated to a heterologous sequence to encode a fusion protein.

A variety of expression vector/host systems, as known in the art, may be utilized to contain and express sequences encoding the proteins or fusion proteins. These include, but are not limited to, micro-organisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus, adenovirus, adeno-associated virus, lentivirus, retrovirus); plant cell systems

transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or PBR322 plasmids); or animal cell systems.

5 The presence of polynucleotide sequences of the invention in a sample can be detected by DNA-DNA or DNA-RNA hybridization and/or amplification using probes or portions or fragments of said polynucleotides. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences specific for the gene to detect transformants
10 containing DNA or RNA encoding the corresponding protein. As used herein 'oligonucleotides' or 'oligomers' refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplier.

15 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting polynucleotide sequences include oligo-labeling, nick translation, end-labeling
20 of RNA probes, PCR amplification using a labeled nucleotide, or enzymatic synthesis. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, Mich.); Promega (Madison Wis.); and U.S. Biochemical Corp., (Cleveland, Ohio).

25 The presence of DG008, DG065, DG210, or DG239 in a sample can be determined by immunological methods or activity measurement. A variety of protocols for detecting and measuring the expression of proteins, using either polyclonal or monoclonal antibodies specific for the protein or reagents for determining protein activity are known in the art. Examples include
30 enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on the protein is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox D.E. et al. (1983; J. Exp. Med. 158:
35 1211-1226).

Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, fluorescent, chemiluminescent or chromogenic agents as well as substrates, co-factors, inhibitors, magnetic particles, and the like.

5

The nucleic acids encoding the proteins of the invention can be used to generate transgenic animal or site-specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination, where the normal locus of the genes encoding the proteins of the invention is altered.

10

Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. The modified cells or animal are useful in the study of the function and regulation of the proteins of the invention. For example, a series of small deletions and/or substitutions may be made in the genes that encode the proteins of the invention to determine the role of particular domains of the protein, functions in pancreatic differentiation, etc.

Specific constructs of interest include anti-sense molecules, which will block the expression of the proteins of the invention, or expression of dominant negative mutations. A detectable marker, such as for example lac-Z, may be introduced in the locus of the genes of the invention, where up-regulation of expression of the genes of the invention will result in an easily detected change in phenotype.

25

One may also provide for expression of the genes of the invention or variants thereof in cells or tissues where it is not normally expressed or at abnormal times of development. In addition, by providing expression of the proteins of the invention in cells in which they are not normally produced, one can induce changes in cell behavior.

30

DNA constructs for homologous recombination will comprise at least portions of the genes of the invention with the desired genetic modification, and will include regions of homology to the target locus. DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and/or negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For non-human embryonic

stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig etc. Such cells are grown on an appropriate fibroblast-feeder layer or grown in presence of leukemia inhibiting factor (LIF).

5 The data disclosed in this Invention show that the DG008, DG065, DG210, or DG239 nucleic acids and proteins and effector/modulator molecules thereof are useful in diagnostic and therapeutic applications implicated, for example, but not limited to, pancreatic diseases (e.g. diabetes mellitus, such as insulin
10 dependent diabetes mellitus and/or non insulin dependent diabetes mellitus), obesity, metabolic syndrome, eating disorder, cachexia, hypertension, coronary heart disease, hypercholesterolemia (dyslipidemia), and/or gallstones. Further, the data show that the DG008, DG065, DG210, or DG239 nucleic acids and proteins and effector/modulator molecules thereof are useful
15 for the modulation, e.g. stimulation of pancreatic development, and/or for the regeneration of pancreatic cells or tissues, e.g. cells having exocrinous functions such as acinar cells, centroacinar cells and/or ductal cells and/or cells having endocrinous functions, particularly cells in Langerhans islets such as alpha-, beta-, delta- and/or PP-cells, more particularly beta-cells. Hence,
20 diagnostic and therapeutic uses for the proteins of the invention nucleic acids and proteins of the invention are, for example but not limited to, the following:
(i) tissue regeneration in vitro and in vivo (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues), (ii) small molecule drug target, (iii) antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) diagnostic and/or prognostic
25 marker, (v) protein therapy, (vi) gene therapy (gene delivery/gene ablation), and / or (vii) research tools.

30 For example, but not limited to, cDNAs encoding the proteins of the invention and particularly their human homologues may be useful in stimulating, enhancing or regulating the regeneration of tissues, and the proteins of the invention and particularly their human homologues may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, pancreatic diseases (e.g. diabetes),
35 obesity, and/or metabolic syndrome as described above.

In one embodiment of the invention, administration of DG008, DG065, DG210, or DG239 nucleic acids and proteins and/or effectors/modulators thereof in a pharmaceutical composition to a subject in need thereof, particularly a human patient lead to an at least partial regeneration of, for example, pancreas cells.

5 The composition will then at least partially restore normal pancreatic function. In one example, these cells are beta cells of the islets which will contribute to the improvement of a diabetic state. With the administration of this composition e.g. on a short term or regular basis, an increase in beta cell mass can be achieved. This effect upon the body reverses the condition of diabetes partially

10 or completely. As the subject's blood sugar level improves, the dosage administered may be reduced in strength. In at least some cases further administration can be discontinued entirely and the subject continues to produce a normal amount of insulin without further treatment. The subject is thereby not only treated but cured entirely of a diabetic condition. However,

15 even moderate improvements in beta cell mass can lead to a reduced requirement for exogenous insulin, improved glycemic control and a subsequent reduction in diabetic complications. In another example, other cells of the pancreas can be regenerated in vivo or in vitro to cure a certain condition. Beside diabetes, the compositions of the present invention will also

20 have efficacy for treatment of patients with other pancreatic diseases such as pancreatic cancer, dysplasia, or pancreatitis.

The DG008, DG065, DG210, or DG239 nucleic acids and proteins and effectors/modulators thereof are useful in diagnostic and therapeutic applications implicated in various embodiments as described below. For example, but not limited to, cDNAs encoding the proteins of the invention and particularly their human homologues may be useful in gene therapy, and the proteins of the invention and particularly their human homologues may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, pancreatic diseases (e.g. diabetes), obesity, and/or metabolic syndrome as described above.

35 The nucleic acids of the invention or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acids or the proteins are to be assessed. Further antibodies that bind immunospecifically to the novel substances of the invention may be used in

therapeutic or diagnostic methods.

For example, in one aspect, antibodies, which are specific for the proteins of the invention and homologous proteins, may be used directly as an effector/modulator, e.g. an antagonist or an agonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein. The antibodies may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralising antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with the protein or any fragment or oligopeptide thereof, which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. It is preferred that the peptides, fragments or oligopeptides used to induce antibodies to the protein have an amino acid sequence consisting of at least five amino acids, and more preferably at least 10 amino acids.

Monoclonal antibodies to the proteins may be prepared using any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler G. and Milstein C., (1975) Nature 256: 495-497; Kozbor D. et al., (1985) J. Immunol. Methods 81: 31-42; Cote R.J. et al., (1983) Proc. Natl. Acad. Sci. 80: 2026-2030; Cole S.P. et al., (1984) Mol. Cell Biol. 62: 109-120).

In addition, techniques developed for the production of 'chimeric antibodies', the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison S.L. et al., (1984) Proc. Natl. Acad. Sci. 81: 6851-6855; Neuberger M.S. et al., (1984) Nature 312: 604-608; Takeda S. et al., (1985) Nature 314: 452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to

produce single chain antibodies specific for the proteins of the invention and homologous proteins. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Kang A.S. et al., (1991) Proc. Natl. Acad. Sci. 88: 11120-11123). Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi R. et al., (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter G. and Milstein C., (1991) Nature 349: 293-299).

10

Antibody fragments, which contain specific binding sites for the proteins may also be generated. For example, such fragments include, but are not limited to, the F(ab')₂ fragments which can be produced by Pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse W.D. et al., (1989) Science 246: 1275-1281).

20

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding and immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering protein epitopes are preferred, but a competitive binding assay may also be employed (Maddox, *supra*).

25

30

In another embodiment of the invention, the polynucleotides or fragments thereof or nucleic acid effector/modulator molecules such as antisense molecules, aptamers, RNAi molecules or ribozymes may be used for therapeutic purposes. In one aspect, aptamers, i.e. nucleic acid molecules, which are capable of binding to a protein of the invention and modulating its activity, may be generated by a screening and selection procedure involving the use of combinatorial nucleic acid libraries.

35

In a further aspect, antisense molecules may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding DG008, DG065, DG210, or DG239 and homologous proteins. Thus,
5 antisense molecules may be used to modulate/effect protein activity or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding the proteins. Expression vectors derived from retroviruses,
10 adenovirus, herpes or vaccinia viruses or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods, which are well known to those skilled in the art, can be used to construct recombinant vectors, which will express antisense molecules complementary to the polynucleotides of the genes encoding the
15 proteins of the invention and homologous proteins. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra). Genes encoding the proteins of the invention and homologous proteins can be turned off by transforming a cell or tissue with expression vectors, which express high levels of polynucleotides that encode the proteins of the invention and homologous proteins or fragments thereof. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the
20 absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.
25

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, e.g. DNA, RNA or nucleic acid analogues such as PNA, to the control regions of the genes encoding DG008, DG065, DG210,
30 or DG239 and homologous proteins, i.e., the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it cause inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee J.E. et al., (1994) Gene 149: 109-114;

Huber B.E. and Carr B.I., (1994) Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.). The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

5

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples, which may be used, include engineered hammerhead motif ribozyme molecules that can be specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding the proteins of the invention and homologous proteins. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

25

Nucleic acid effector/modulator molecules, e.g. antisense molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences. Such DNA sequences may be incorporated into a variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells or tissues. RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or modifications in the nucleobase, sugar and/or phosphate moieties, e.g. the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production

30

35

of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods, which are well known in the art. Any of the therapeutic methods described above may be applied to any suitable subject including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of DG008, DG065, DG210, or DG239 nucleic acids and the proteins and homologous nucleic acids or proteins, antibodies to the proteins of the invention and homologous proteins, mimetics, agonists, antagonists or inhibitors of the proteins of the invention and homologous proteins or nucleic acids. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone or in combination with other agents, drugs or hormones. The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active compounds into

preparations, which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.).

- 5 Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compounds, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of preadipocyte cell lines or in animal models, usually mice, rabbits, dogs or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. A therapeutically effective dose refers to that amount of active ingredient, for example the DG008, DG065, DG210, or DG239 nucleic acids or proteins or fragments thereof or antibodies, which is sufficient for treating a specific condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions, which exhibit large therapeutic indices, are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use.
- 10 The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration. The exact dosage will be determined by the practitioner, in light of factors related to the subject that
- 15 requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors, which may be taken into account, include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities,
- 20 and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week or once every two weeks depending on half-life and clearance rate of the particular formulation. Normal
- 25
- 30
- 35

dosage amounts may vary from 0.1 to 100,000 microg, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

In another embodiment, antibodies which specifically bind to the proteins may be used for the diagnosis of conditions or diseases characterized by or associated with over- or under-expression of the proteins of the invention and homologous proteins or in assays to monitor patients being treated with the proteins of the invention and homologous proteins, or effectors/modulators thereof, e.g. agonists, antagonists, or inhibitors. Diagnostic assays include methods which utilize the antibody and a label to detect the protein in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules, which are known in the art may be used several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring proteins are known in the art and provide a basis for diagnosing altered or abnormal levels of gene expression. Normal or standard values for gene expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibodies to the protein under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of protein expressed in control and disease, samples e.g. from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides specific for the DG008, DG065, DG210, or DG239 proteins and homologous proteins may be used for diagnostic purposes. The polynucleotides, which may be used, include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene

expression in biopsied tissues in which gene expression may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess gene expression, and to monitor regulation of protein levels during therapeutic intervention.

5

In one aspect, hybridization with probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding the proteins of the invention and homologous proteins or closely related molecules, may be used to identify nucleic acid sequences which encode the respective protein. The hybridization probes of the subject invention may be DNA or RNA and are preferably derived from the nucleotide sequence of the polynucleotide encoding the proteins of the invention or from a genomic sequence including promoter, enhancer elements, and introns of the naturally occurring gene. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as ^{32}P or ^{35}S or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

10

15

Polynucleotide sequences specific for DG008, DG065, DG210, or DG239 proteins and homologous nucleic acids may be used for the diagnosis of conditions or diseases, which are associated with the expression of the proteins. Examples of such diseases include the pancreatic diseases (e.g. diabetes), obesity, metabolic syndrome, and/or others. Polynucleotide sequences specific for the DG008, DG065, DG210, or DG239 proteins and homologous proteins may also be used to monitor the progress of patients receiving treatment for pancreatic diseases (e.g. diabetes), obesity, and/or metabolic syndrome. The polynucleotide sequences may be used qualitative or quantitative assays, e.g. in Southern or Northern analysis, dot blot or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered gene expression.

20

25

30

In a particular aspect, the DG008, DG065, DG210, or DG239 nucleotide sequences may be useful in assays that detect activation or induction of various metabolic diseases or dysfunctions. The nucleotide sequences may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes.

After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. The presence of altered levels of nucleotide sequences encoding the proteins of the invention and homologous proteins in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disease associated with expression of the DG008, DG065, DG210, or DG239 proteins and homologous proteins, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence or a fragment thereof, which is specific for the nucleic acids encoding the proteins of the invention and homologous nucleic acids, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease. Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that, which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to pancreatic diseases (e.g. diabetes), obesity, and/or metabolic syndrome, the presence of an unusual amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the metabolic diseases and disorders.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding the proteins of the invention and homologous proteins may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically or produced from a recombinant source. Oligomers will 5 preferably consist of two nucleotide sequences, one with sense orientation (5prime.fwdarw.3prime) and another with antisense (3prime.rarw.5prime), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers or even a degenerate pool of oligomers may be employed under less stringent conditions 10 for detection and/or quantification of closely related DNA or RNA sequences.

In another embodiment of the invention, the nucleic acid sequences may also be used to generate hybridization probes, which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a 15 particular chromosome or to a specific region of the chromosome using well known techniques. Such techniques include FISH, FACS or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price C.M., (1993) Blood Rev. 7: 127-134, and Trask 20 B.J., (1991) Trends Genet. 7: 149-154. FISH (as described in Verma R.S. and Babu A., (1989) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, N.Y.). The results may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of 25 genetic map data can be found in the 1994 Genome Issue of Science (265: 1981f). Correlation between the location of the gene encoding the proteins of the invention on a physical chromosomal map and a specific disease or predisposition to a specific disease, may help to delimit the region of DNA associated with that genetic disease.

30 The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals. An analysis of polymorphisms, e.g. single nucleotide polymorphisms may be carried out. Further, in situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established 35 chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a

particular human chromosome is not known. New sequences can be assigned to chromosomal arms or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti R.A. et al., (1988) Nature 336: 577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequences of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals.

In another embodiment of the invention, the proteins of the invention, their catalytic or immunogenic fragments or oligopeptides thereof, an in vitro model, a genetically altered cell or animal, can be used for screening libraries of compounds in any of a variety of drug screening techniques. One can identify effectors, e.g. receptors, enzymes, proteins, ligands, or substrates that bind to, modulate or mimic the action of one or more of the DG008, DG065, DG210, or DG239 proteins of the invention. The protein or fragment thereof employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between the DG008, DG065, DG210, or DG239 proteins of the invention and the agent tested, may be measured. Agents could also, either directly or indirectly, influence the activity of the proteins of the invention.

In addition activity of the proteins of the invention against their physiological substrate(s) or derivatives thereof could be measured in cell-based or cell-free assays. Agents may also interfere with posttranslational modifications of the protein, such as phosphorylation and dephosphorylation, farnesylation, palmitoylation, acetylation, alkylation, ubiquitination, proteolytic processing, subcellular localization and degradation. Moreover, agents could influence the dimerization or oligomerization of the proteins of the invention or, in a heterologous manner, of the proteins of the invention with other proteins, for example, but not exclusively, docking proteins, enzymes, receptors, or translation factors. Agents could also act on the physical interaction of the proteins of this invention with other proteins, which are required for protein function, for example, but not exclusively, their downstream signaling.

Methods for determining protein-protein interaction are well known in the art. For example binding of a fluorescently labeled peptide derived from the interacting protein to the DG008, DG065, DG210, or DG239 protein of the invention, or vice versa, could be detected by a change in polarisation. In case that both binding partners, which can be either the full length proteins as well as one binding partner as the full length protein and the other just represented as a peptide are fluorescently labeled, binding could be detected by fluorescence energy transfer (FRET) from one fluorophore to the other. In addition, a variety of commercially available assay principles suitable for detection of protein-protein interaction are well known in the art, for example but not exclusively AlphaScreen (PerkinElmer) or Scintillation Proximity Assays (SPA) by Amersham. Alternatively, the interaction of the DG008, DG065, DG210, or DG239 proteins of the invention with cellular proteins could be the basis for a cell-based screening assay, in which both proteins are fluorescently labeled and interaction of both proteins is detected by analysing cotranslocation of both proteins with a cellular imaging reader, as has been developed for example, but not exclusively, by Cellomics or EvotecOAI. In all cases the two or more binding partners can be different proteins with one being the protein of the invention, or in case of dimerization and/or oligomerization the protein of the invention itself.

Of particular interest are screening assays for agents that have a low toxicity for mammalian cells. The term "agent" as used herein describes any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking the physiological function of one or more of the proteins of the invention. Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 Daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise carbocyclic or heterocyclic structures and/or aromatic or polycyclic aromatic structures substituted with one or more of the above functional groups.

Candidate agents are also found among biomolecules including peptides,

saccharides, fatty acids, steroids, purines, pyrimidines, nucleic acids and derivatives, structural analogs or combinations thereof. Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs. Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal.

Another technique for drug screening, which may be used, provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to the proteins of the invention large numbers of different small test compounds, e.g. aptamers, peptides, low-molecular weight compounds etc., are provided or synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with the proteins or fragments thereof, and washed. Bound proteins are then detected by methods well known in the art. Purified proteins can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support. In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound for binding the protein. In this manner, the antibodies can be used to detect the presence of any peptide, which shares one or more antigenic determinants with the protein.

Compounds that bind DG008, DG065, DG210, or DG239 proteins, e.g. antibodies, are useful for the identification or enrichment of cells, which are positive for the expression of the proteins of the invention, from complex cell

mixtures. Such cell populations are useful in transplantation, for experimental evaluation, and as source of lineage and cell specific products, including mRNA species useful in identifying genes specifically expressed in these cells, and as target for the identification of factors or molecules that can affect them.

5 Cells expressing the protein of the invention or which have been treated with the protein of the invention are useful in transplantation to provide a recipient with pancreatic islet cells, including insulin producing beta cells; for drug screening; experimental models of islet differentiation and interaction with other cell types; in vitro screening assays to define growth and differentiation
10 factors, and to additionally characterize genes involved in islet development and regulation; and the like. The native cells may be used for these purposes, or they may be genetically modified to provide altered capabilities. Cells from a regenerating pancreas, from embryonic foregut, stomach and duodenum, or other sources of pancreatic progenitor cells may be used as a starting
15 population. The progenitor cells may be obtained from any mammalian species, e.g. equine, bovine, porcine, canine, feline, rodent, e.g. mice, rats, hamster, primate, etc. particularly human.

In another embodiment, in a high-throughput screening method, the cells are
20 transfected with a DNA construct, e.g. a viral or non-viral vector containing a reporter gene, e.g. the lacZ gene or the GFP gene, under regulatory control of a promoter of a gene involved in for example beta-cell differentiation, e.g. a promoter of a gene stimulation beta-cell differentiation, preferably a Pax4 promoter. The transfected cells are divided into aliquots and each aliquot is contacted with a test substance, e.g., candidate 1 candidate 2, candidate 3, etc. The activity of the reporter gene corresponds to the capability of the test compound to induce beta-cell differentiation.

In a further embodiment, which may be combined with the high-throughput screening as described above, a medium throughput validation is carried out.
30 Therein, the test compound is added to stem cells being cultivated and the insulin production is determined. Following an initial high throughput assay, such as the cell based assay outlined above where for example a Pax4 promoter is used as marker for beta-cell regeneration, the activity of candidate molecules to induce beta-cell differentiation is tested in a validation assay comprising adding said compounds to the culture media of the embryoid bodies. Differentiation into insulin-producing cells is then evaluated, e.g. by
35

comparison to wild type and/or Pax4 expressing ES cells to assess the effectiveness of a compound.

5 The nucleic acids encoding the DG008, DG065, DG210, or DG239 proteins of the invention can be used to generate transgenic cell lines and animals. These transgenic non-human animals are useful in the study of the function and regulation of the proteins of the invention *in vivo*. Transgenic animals, particularly mammalian transgenic animals, can serve as a model system for the investigation of many developmental and cellular processes common to 10 humans. A variety of non-human models of metabolic disorders can be used to test modulators of the protein of the invention. Misexpression (for example, overexpression or lack of expression) of the protein of the invention, particular feeding conditions, and/or administration of biologically active compounds can create models of metabolic disorders.

15 In one embodiment of the invention, such assays use mouse models of insulin resistance and/or diabetes, such as mice carrying gene knockouts in the leptin pathway (for example, ob (leptin) or db (leptin receptor) mice), as described above. In addition to testing the expression of the proteins of the invention in 20 such mouse strains (see Examples), these mice could be used to test whether administration of a candidate modulator alters for example lipid accumulation in the liver, in plasma, or adipose tissues using standard assays well known in the art, such as FPLC, colorimetric assays, blood glucose level tests, insulin tolerance tests and others.

25 Transgenic animals may be made through homologous recombination in embryonic stem cells, where the normal locus of the gene encoding the protein of the invention is mutated. Alternatively, a nucleic acid construct encoding the protein is injected into oocytes and is randomly integrated into the genome. 30 One may also express the genes of the invention or variants thereof in tissues where they are not normally expressed or at abnormal times of development. Furthermore, variants of the genes of the invention like specific constructs expressing anti-sense molecules or expression of dominant negative mutations, which will block or alter the expression of the proteins of the invention may be randomly integrated into the genome. A detectable marker, 35 such as lac Z or luciferase may be introduced into the locus of the genes of the invention, where up-regulation of expression of the genes of the invention will

result in an easily detectable change in phenotype. Vectors for stable integration include plasmids, retroviruses and other animal viruses, yeast artificial chromosomes (YACs), and the like. DNA constructs for homologous recombination will contain at least portions of the genes of the invention with the desired genetic modification, and will include regions of homology to the target locus. Conveniently, markers for positive and negative selection are included. DNA constructs for random integration do not need to contain regions of homology to mediate recombination. DNA constructs for random integration will consist of the nucleic acids encoding the proteins of the invention, a regulatory element (promoter), an intron and a polyadenylation signal. Methods for generating cells having targeted gene modifications through homologous recombination are known in the field. For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig, etc. Such cells are grown on an appropriate fibroblast-feeder layer and are grown in the presence of leukemia inhibiting factor (LIF). ES or embryonic cells may be transfected and can then be used to produce transgenic animals. After transfection, the ES cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be selected by employing a selection medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination. Colonies that are positive may then be used for embryo manipulation and morula aggregation. Briefly, morulae are obtained from 4 to 6 week old superovulated females, the Zona Pellucida is removed and the morulae are put into small depressions of a tissue culture dish. The ES cells are trypsinized, and the modified cells are placed into the depression closely to the morulae. On the following day the aggregates are transferred into the uterine horns of pseudopregnant females. Females are then allowed to go to term. Chimeric offsprings can be readily detected by a change in coat color and are subsequently screened for the transmission of the mutation into the next generation (F1-generation). Offspring of the F1-generation are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogenic or congenic grafts or transplants, or in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animal, domestic animals, etc., for example, mouse, rat, guinea pig, sheep, cow, pig, and others. The transgenic animals

may be used in functional studies, drug screening, and other applications and are useful in the study of the function and regulation of the proteins of the invention *in vivo*.

- 5 Finally, the invention also relates to a kit comprising at least one of
- (a) a nucleic acid molecule coding for a protein of the invention or a functional fragment thereof;
 - (b) a protein of the invention or a fragment or an isoform thereof;
 - (c) a vector comprising the nucleic acid of (a);
 - 10 (d) a host cell comprising the nucleic acid of (a) or the vector of (c);
 - (e) a polypeptide encoded by the nucleic acid of (a);
 - (f) a fusion polypeptide encoded by the nucleic acid of (a);
 - (g) an antibody, an aptamer or another effector/modulator against the nucleic acid of (a) or the polypeptide of (b), (e) or (f) and
 - 15 (h) an anti-sense oligonucleotide of the nucleic acid of (a).

The kit may be used for diagnostic or therapeutic purposes or for screening applications as described above. The kit may further contain user instructions.

20

The Figures show:

Fig. 1 shows human DG008 nucleic acid and protein.

Fig. 1A shows the nucleic acid sequence encoding human DG008 protein (SEQ ID NO: 1; GenBank Accession Number NM_004684).

Fig. 1B shows the amino acid sequence (one-letter code) of human DG008 (SEQ ID NO: 2; GenBank Accession Number NP_004675).

Fig. 2 shows the quantitative (real-time PCR) analysis of DG008 protein expression in mammalian tissues and cells.

Fig. 2A shows the analysis of DG008 expression in wild-type mouse tissues.

Fig. 2B shows the analysis of DG008 expression in genetically obese mice (ob/ob-mice) and fasted mice (fasted-mice) compared to wild-type mice (wt-mice).

Fig. 2C shows the analysis of DG008 expression in mice fed with a high fat diet compared to mice fed with a control diet.

Fig. 2D shows the analysis of DG008 expression in mouse tissues from non-obese-diabetic (NOD) mice compared to wild-type mice.

Fig. 2E shows the analysis of DG008 expression in mammalian fibroblast (3T3-L1) cells, during the differentiation from preadipocytes to mature adipocytes.

Fig. 3 shows human DG065 nucleic acid and protein.

Fig. 3A shows the nucleic acid sequence encoding human DG065 protein (SEQ ID NO: 3; GenBank Accession Number NM_022138).

Fig. 3B shows the amino acid sequence (one-letter code) of human DG065 protein (SEQ ID NO: 4; GenBank Accession Number NP_071421).

Fig. 4 shows the analysis of DG065 protein expression in mammalian tissues and cells.

Fig. 4A shows the quantitative (real-time PCR) analysis of DG065 expression in wild-type mouse tissues.

Fig. 4B shows the quantitative (real-time PCR) analysis of DG065 expression in genetically obese mice (ob/ob-mice) and fasted mice (fasted-mice) compared to wild-type mice (wt-mice).

Fig. 4C shows the quantitative (real-time PCR) analysis of DG065 expression in mice fed with a high fat diet compared to mice fed with a control diet.

Fig. 4D shows the quantitative (real-time PCR) analysis of DG065 expression in mouse tissues from non-obese-diabetic (NOD) mice compared to wild-type mice.

Fig. 4E shows the quantitative (real-time PCR) analysis of DG065 expression in mammalian fibroblast (3T3-L1) cells, during the differentiation from preadipocytes to mature adipocytes.

Fig. 4F shows the microarray analysis of DG065 expression in mammalian fibroblast (3T3-L1) cells, during the differentiation from preadipocytes to mature adipocytes.

Fig. 4G shows the microarray analysis of DG065 expression in human abdominal adipocyte cells, during the differentiation from preadipocytes to mature adipocytes.

Fig. 5 shows human DG210 nucleic acid and protein.

Fig. 5A shows the nucleic acid sequence of human DG210 protein (SEQ ID

NO: 5; GenBank Accession Number NM_003012).

Fig. 5B shows the amino acid sequence (one-letter code) of human DG210 protein (SEQ ID NO: 6; GenBank Accession Number NP_003003).

5 Fig. 6 shows the quantitative (real-time PCR) analysis of DG210 protein expression in mammalian tissues.

Fig. 6A shows the analysis of DG210 expression in wild-type mouse tissues.

10 Fig. 6B shows the analysis of DG210 expression in genetically obese mice (ob/ob-mice) and fasted mice (fasted-mice) compared to wild-type mice (wt-mice).

Fig. 6C shows the analysis of DG210 expression in mice tissues from non-obese-diabetic (NOD) mice compared to wild-type mice.

Fig. 7 shows human DG239 nucleic acid and protein.

15 Fig. 7A shows the nucleic acid sequence encoding the human DG239 protein (SEQ ID NO: 7; GenBank Accession Number NM_002087).

Fig. 7B shows the amino acid sequence (one-letter code) of human DG239 protein (SEQ ID NO: 8; GenBank Accession Number NP_002078).

20 Fig. 8 shows the quantitative (real-time PCR) analysis of DG239 protein expression in mammalian tissues and cells.

Fig. 8A shows the analysis of DG239 expression in wild-type mouse tissues.

25 Fig. 8B shows the analysis of DG239 expression in genetically obese mice (ob/ob-mice) and fasted mice (fasted-mice) compared to wild-type mice (wt-mice).

Fig. 8C shows the analysis of DG239 expression in mice fed with a high fat diet compared to mice fed with a control diet.

Fig. 8D shows the analysis of DG239 expression in mouse tissues from non-obese-diabetic (NOD) mice compared to wildtype mice.

30 Fig. 8E shows the analysis of DG239 expression in mammalian fibroblast (3T3-L1) cells, during the differentiation from preadipocytes to mature adipocytes.

35 The examples illustrate the invention:

Example 1: Identification of secreted factors expressed in pancreas

A screen for secreted factors expressed in developing mouse pancreas was carried out according to methods known by those skilled in the art (see, for example Pera E.M. and De Robertis E.M., (2000) *Mech Dev* 96: 183-195) with several modifications.

Expression cDNA library:

During organogenesis, the pancreatic bud is surrounded and influenced by the associated mesenchyme (see for example, Madsen O.D. et al., (1996) *Eur. J. Biochem.* 242: 435-445 and Slack, J.M., (1995) *Development* 121: 1569-1580). Recently, it was suggested, that white adipocytes origin directly from mesenchymal cells (Atanossova P.K., (2003) *Folia Med.* 45: 41-45). During embryogenesis, the innervation and vascularization of the pancreas can be observed. Therefore, the tissue used in the screen might have contained besides pancreatic cells some adipocyte precursors, blood vessels, as well as neuronal cells.

A mouse embryonic stage 9.5-15 pancreatic bud library was prepared in pCMVSPORT-6 vector using SUPERSCRIPT Plasmid System from Invitrogen according to the manufacturer's instructions. The non-amplified library was electroporated into MaxEff DH10B cells (Invitrogen).

Secretion cloning

Bacterial clones were picked with sterile toothpicks from agar plates and cultured in 96-deep-well microtiter plates in LB-ampicillin (see Sambrook et al., supra). Aliquots of 8 cultures were pooled, and plasmid DNA was isolated using the BioRobot_9600 apparatus according to the manufacturer's instructions (Qiagen; QIAprep(r) Turbo BioRobot Kit. Human 293 cell culture cells were cultured in 75 ml tissue culture flasks in DMEM and 10% fetal calf serum. At 90-99% confluence, the cells were splitted at 1:3 ratio and plated onto poly-D-lysine (Sigma) coated 96-well plates. Cells were transfected with 100-500 ng plasmid using lipofectamine 2000 (Invitrogen). After 6 hours, the medium was exchanged for fresh complete growth medium. 24 hours after transfection, the cells were washed twice with DMEM without cysteine and methionine (Invitrogen), supplemented with 1% dialysed Bovine serum (Sigma) with 50 microgram per ml Heparin (Sigma) and glutamine. The cells were

labeled radioactively ('S35 Met-label', from Hartmann Analytic GmbH). After 12 hours, aliquots of the supernatants were harvested in 96-well PCR plates and subjected to SDS gel electrophoresis in precast 4(20% gradient polyacrylamide Criterion gels (Biorad) under reducing conditions, using Criterion Dodeca Cell gel running chamber (Biorad). The gels were fixed in 10% acetic acid, 25% isopropanol for 30 min, soaked 15-30 min in AMPLIFY reagent (Amersham), dried and exposed to X-OMAT (AR) film (Kodak). Positive clones were identified and regrown in 96-well-plates. DNA of individual clones was prepared and used for transfection as described above. If one of the clones yielded proteins of the same size as that of the original pool, a positive clone was identified. Positive clones were partially sequenced from the 5' end (SEQLAB, Goettingen).

15 **Example 2: Identification of the human DG008, DG065, DG210, and DG239 homologous nucleic acid and protein sequence**

The term "polynucleotide comprising the nucleotide sequence as shown in GenBank Accession number" relates to the expressible gene of the nucleotide sequences deposited under the corresponding GenBank Accession number. The term "GenBank Accession number" relates to NCBI GenBank database entries (Ref.: Benson D.A. et al., (2000) Nucleic Acids Res. 28: 15-18).

DG008, DG065, DG210, or DG239 homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are nucleic acids comprising human DG008, DG065, DG210, or DG239 homologs. The following mouse sequences were identified in the 'secreted factor screen': GenBank Accession Numbers NM_010097 and NP_034227 (DG008), GenBank Accession Numbers NM_022315 and NP_071710 (DG065), GenBank Accession Numbers NM_013834 and NP_038862 (DG210), and GenBank Accession Numbers NM_008175 and NP_032201 (DG239).

Sequences homologous mouse DG008, DG065, DG210, or DG239 were identified using the publicly available program BLASTP 2.2.3 of the non-redundant protein data base of the National Center for Biotechnology Information (NCBI) (see, Altschul S.F. et al., (1997) Nucleic Acids Res. 25:

3389-3402). The best human homolog of mouse DG008 is GenBank Accession Number NM_004684 (SEQ ID NO: 1; see Fig. 1A) and GenBank Accession Number NP_004675 (SEQ ID NO: 2; see Fig. 1B), the best human homolog of mouse DG065 is GenBank Accession Number NM_022138 (SEQ ID NO: 3; see Fig. 3A) and GenBank Accession Number NP_071421 (SEQ ID NO: 4; see Fig. 3B), the best human homolog of mouse DG210 is GenBank Accession Number NM_003012 (SEQ ID NO: 5; see Fig. 5A) and GenBank Accession NP_003003 (SEQ ID NO: 6; see Fig. 5B), and the best homolog of mouse DG239 is GenBank Accession Number NM_002087 (SEQ ID NO: 7; see Fig. 7A) and GenBank Accession Number NP_002078 (SEQ ID NO: 8; see Fig. 7B).

Example 3: Quantitative analysis of the expression of the nucleic acids of the invention in mammalian (mouse) tissues

To analyse the expression of the DG008, DG065, DG210, or DG239 mRNA disclosed in this invention in mammalian tissues, several mouse strains (preferably mice strains C57Bl/6J, C57Bl/6 ob/ob, C57Bl/KS db/db, and Non-Obese-Diabetic (NOD) mice, which are standard model systems in obesity and diabetes research) were purchased from Harlan Winkelmann (33178 Borch, Germany) and Taconic M & B (Germantown, NY 12526, U.S.A.), respectively, and maintained under constant temperature (preferably 22°C), 40% humidity and a light / dark cycle of preferably 14 / 10 hours. The mice were fed a standard chow (for example, from ssniff Spezialitäten GmbH, order number ssniff M-Z V1126-000). For the fasting experiment ("fasted wild type mice"), wild type mice were starved for 48 h without food, but only water supplied ad libitum (see, for example, Schnetzler B. et al., (1993) J Clin Invest 92: 272-280, Mizuno T.M. et al., (1996) Proc Natl Acad Sci 93: 3434-3438). In a further experiment wild-type (wt) mice were fed a control diet (preferably Altromin C1057 mod control, 4.5% crude fat) or high fat diet (preferably Altromin C1057 mod. high fat, 23.5% crude fat). Animals were sacrificed at an age of 6 to 8 weeks. The animal tissues were isolated according to standard procedures known to those skilled in the art, snap frozen in liquid nitrogen and stored at -80°C until needed.

For analyzing the role of the proteins disclosed in this invention in the in vitro

differentiation of mammalian cell culture cells for the conversion of pre-adipocytes to adipocytes, mammalian fibroblast (3T3-L1) cells (e.g., Green H. and Kehinde O., (1974) Cell 1: 113-116) were obtained from the American Tissue Culture Collection (ATCC, Hanassas, VA, USA; ATCC- CL 173).
5 3T3-L1 cells were maintained as fibroblasts and differentiated into adipocytes as described in the prior art (e.g., Qiu Z. et al., (2001) J. Biol. Chem. 276: 11988-11995; Slieker L.J. et al., (1998) BBRC 251: 225-229). In brief, cells were plated in DMEM/10% FCS (Invitrogen, Karlsruhe, Germany) at 50,000 cells/well in duplicates in 6-well plastic dishes and cultured in a humidified
10 atmosphere of 5% CO₂ at 37°C. At confluence (defined as day 0: d0) cells were transferred to serum-free (SF) medium, containing DMEM/HamF12 (3:1; Invitrogen), fetuin (300 µg/ml; Sigma, Munich, Germany), transferrin (2 µg/ml; Sigma), pantothenate (17 µM; Sigma), biotin (1 µM; Sigma), and EGF (0.8 nM; Hoffmann-La Roche, Basel, Switzerland). Differentiation was induced by
15 adding dexamethasone (DEX; 1 µM; Sigma), 3-methyl-isobutyl-1-methylxanthine (MIX; 0.5 mM; Sigma), and bovine insulin (5 µg/ml; Invitrogen). Four days after confluence (d4), cells were kept in SF medium, containing bovine insulin (5 µg/ml) until differentiation was completed. At various time points of the differentiation procedure, beginning with day 0 (day
20 of confluence) and day 2 (hormone addition; for example, dexamethasone and 3-isobutyl-1-methylxanthine), up to 10 days of differentiation, suitable aliquots of cells were taken every two days.

RNA was isolated from mouse tissues or cell culture cells using Trizol Reagent (for example, from Invitrogen, Karlsruhe, Germany) and further purified with the RNeasy Kit (for example, from Qiagen, Germany) in combination with an DNase-treatment according to the instructions of the manufacturers and as known to those skilled in the art. Total RNA was reverse transcribed (preferably using Superscript II RNaseH- Reverse Transcriptase, from Invitrogen, Karlsruhe, Germany) and subjected to Taqman analysis preferably using the Taqman 2xPCR Master Mix (from Applied Biosystems, Weiterstadt, Germany; the Mix contains according to the Manufacturer for example AmpliTaq Gold DNA Polymerase, AmpErase UNG, dNTPs with dUTP, passive reference Rox and optimized buffer components) on a GeneAmp 5700
25 Sequence Detection System (from Applied Biosystems, Weiterstadt, Germany).
30
35

The following primer/probe pairs were used for the TaqMan analysis (GenBank Accession Number NM_010097 for the mouse DG008 sequence):

Mouse DG008 forward primer (Seq ID NO: 9): 5'-CCC GGA GAA CGG CGA T-3'; mouse DG008 reverse primer (Seq ID NO: 10): 5'-GTT TGC GCG TTG 5 GTT AGG A-3'; mouse DG008 Taqman probe (Seq ID NO: 11): (5/6-FAM)-CGC GAC CGT GTC CAC TTC CTA TGT G- (5/6-TAMRA).

The following primer/probe pairs were used for the TaqMan analysis (GenBank Accession Number NM_022315 for the mouse DG065 sequence):

10 Mouse DG065 forward primer (Seq ID NO: 12): 5`- GGA TAC TGC TGG TGT GTG CTA GTG -3'; mouse DG065 reverse primer (Seq ID NO: 13): 5`- GGT TGC TCA TAC CTT GTG GAG G-3'; mouse DG065 Taqman probe (Seq ID NO: 14): (5/6-FAM)- ACA CTG GAC GGC CCA TTC CTG G -(5/6-TAMRA).

15 The following primer/probe pairs were used for the TaqMan analysis (GenBank Accession Number NM_013834 for the mouse DG210 sequence):

20 Mouse DG210 forward primer (SEQ ID NO: 15):5'-CGA GCC GGT CAT GCA GTT-3'; mouse DG210 reverse primer (SEQ ID NO: 16): 5'-GAT GCA GAC GTC GCC CTC-3'; mouse DG210 Taqman probe (SEQ ID NO: 17): (5/6-FAM)-ACT GGC CCG AGA TGC TCA AAT GTG AC-(5/6-TAMRA).

25 The following primer/probe pairs were used for the TaqMan analysis (GenBank Accession Number NM_008175 for the mouse DG239 sequence): Mouse DG239 forward primer (Seq ID NO: 18): 5'- TTC ACA CAC GAT GCG TTT CAC -3'; mouse DG239 reverse primer (Seq ID NO: 19): 5'- CCT GTT GGT CTT TTG TGC AGG -3'; mouse DG239 Taqman probe (Seq ID NO: 20): (5/6-FAM)- ACG GGC ACC CAC ACC CTA CTA AAG AAG TT -(5/6-TAMRA).

30 In Fig. 2, 4, 6, and 8, the relative RNA-expression is shown on the Y-axis, in Fig. 2A to 2, 4A to 4D, 6A to 6C, and 8A to 8 the tissues tested are given on the X-axis. WAT refers to white adipose tissue, BAT refers to brown adipose tissue. In Fig. 2, 4E to 4G, and 8 the X-axis represents the time axis. 'd0' refers to day 0 (start of the experiment), 'd2' - 'd12' refers to day 2 - day 1 of 35 adipocyte differentiation.

Taqman analysis revealed that DG008 is expressed in several mammalian tissues, with highest expression levels in hypothalamus and brain in wild type mice. In addition, DG008 is also highly expressed in metabolic active tissue such as white adipose tissue (WAT) and brown adipose tissue (BAT) as depicted in Fig. 2A.

Taqman analysis revealed that DG065 is expressed in several mammalian tissues, with highest expression levels in muscle and spleen of wild type mice. In addition, DG065 is also highly expressed in heart and metabolic active tissues such as white adipose tissue (WAT), brown adipose tissue (BAT) and kidney as depicted in Figure 4A.

Taqman analysis revealed that DG210 is expressed in several mammalian tissues, with highest expression levels in kidney and white adipose tissue (WAT) in wild type mice. In addition, DG210 is also expressed in brown adipose tissue (BAT), and in muscle, brain and hypothalamus as depicted in Fig. 6A.

Taqman analysis revealed that DG239 is expressed in several mammalian tissues, with highest expression levels in white adipose tissue (WAT) of wild type mice. In addition, DG239 is also highly expressed in further metabolic active tissues such as brown adipose tissue (BAT) and liver and at lower but robust levels in muscle as depicted in Fig. 8A.

Further, mouse models of insulin resistance and/or diabetes were used, such as mice carrying gene knockouts in the leptin pathway (for example, ob/ob (leptin) or db/db (leptin receptor/ligand) mice) to study the expression of DG008, DG065, DG210, and DG239. Such mice develop typical symptoms of diabetes, show hepatic lipid accumulation and frequently have increased plasma lipid levels (see Bruning J.C. et al., (1998) Mol. Cell. 2: 559-569).

DG008 is up-regulated in metabolic active tissue (BAT) in genetically induced obese mice (ob/ob) compared to wild type mice (see Fig. 2B).

We also found that DG065 is significantly up-regulated in metabolic active tissues (WAT and BAT) of genetically induced obese mice (ob/ob) compared to wild type mice (see Figure 4B). Further, we found that the expression of DG065 is significantly down-regulated in BAT of starved wild type mice compared to wild type mice fed a standard diet (see Fig. 4B).

We found, that the expression of DG210 is up-regulated in metabolic active tissues, like BAT, muscle and liver of genetically induced obese mice (ob/ob) (compared to wild type mice), and down-regulated in hypothalamus and small intestine of fasted wild type mice (see Fig. 6B).

- 5 We also found that the expression of DG239 is significantly up-regulated in metabolic active tissues (BAT) and muscle in genetically induced obese mice (ob/ob) compared to wild type mice (see Fig. 8B).

Expression of DG008, DG065, and DG239 mRNA was also examined in
10 susceptible wild type mice (for example, C57Bl/6) that show symptoms of diabetes, lipid accumulation, and high plasma lipid levels, if fed a high fat diet. In those mice, the expression of DG008 is up-regulated in WAT supporting that DG008 is involved in the regulation of mammalian metabolism (see Fig. 2C). In susceptible wild type mice, the expression of DG065 is down-regulated
15 in kidney supporting that DG065 is involved in the regulation of mammalian metabolism (see Fig. 4C). In those mice, the expression of DG239 is significantly up-regulated in WAT supporting that DG239 is involved in the regulation of mammalian metabolism (see Fig. 8C).

20 Expression of DG008, DG065, DG210, and DG239 mRNA was also examined non-obese-diabetic (NOD) mice. In those mice, the expression of DG008 is up-regulated in metabolic active tissues e.g. liver and kidney compared to wild type mice, as shown in Fig. 2D.

In NOD mice, expression of DG065 is significantly up-regulated in BAT supporting that DG065 is involved in the regulation of mammalian metabolism (see Fig. 4D).

Expression of DG210 is significantly up-regulated in BAT, colon, brain and liver and down-regulated in small intestine of NOD mice, as shown in Fig. 6C, supporting that DG210 is involved in the regulation of mammalian metabolism.

30 In NOD mice, expression of DG239 is down-regulated in pancreas compared to wild-type mice, as shown in Fig. 8D.

Furthermore, it is shown that the DG008 (see Fig. 2E) and DG065 (see Fig. 4E) mRNA is expressed and regulated during the differentiation into mature
35 adipocytes. We further show that the DG239 (Fig. 8E) mRNA is increased during differentiation of preadipocytes into mature adipocytes. Based on the results in this Example DG008, DG065, DG210 and DG239 proteins might

play an essential role in metabolism, e.g. in adipogenesis or other metabolic disorders.

Example 4: Analysis of the differential expression of transcripts of the proteins of the invention in human tissues

RNA preparation from mouse fibroblast (3T3-L1) cells and human primary adipose tissues was done as described in Example 3. The target preparation, hybridization, and scanning was performed as described in the manufacturers manual (see Affymetrix Technical Manual, 2002, obtained from Affmetrix, Santa Clara, USA).

The expression analysis (using Affymetrix GeneChips) of the DG065 gene using mammalian fibroblast (3T3-L1) and primary human abdominal adipocyte differentiation clearly shows differential expression of mouse and human DG065 in adipocytes. Several independent experiments were done. The experiments show that the DG065 transcripts are most abundant at day 10/12 compared to day 0 during differentiation (see Fig. 4F and 4G).

Thus, the DG065 protein has to be increased in order for the preadipocytes to differentiate into mature adipocytes. The DG065 protein in preadipocytes has the potential to enhance adipose differentiation. Therefore, the DG065 protein might play an essential role in the regulation of human metabolism, in particular in the regulation of adipogenesis and thus it might be an essential role in pancreatic diseases (e.g. diabetes), obesity, and/or metabolic syndrome.

For the purpose of the present invention, it will be understood by the person having average skill in the art, that any combination of any feature mentioned throughout the specification is explicitly disclosed herewith.